Globus Futures

The Globus Team

the globus project™
www.globus.org

Globus 2.0 Release

- Major features
  - MDS 2.1
  - Gram 1.5
  - Packaging tools
  - Data Grid Tools
- New packaging model allows decoupled updates
- Currently in Beta release
  - Complete release
- Will release when confident

U.S. GRIDS Center

- GRIDS = Grid Research, Integration, Deployment, & Support
- NSF-funded center to provide
  - State-of-the-art middleware infrastructure to support national-scale collaborative science and engineering
  - Integration platform for experimental middleware technologies
- ISI, NCSA, SDSC, UC, UW + commercial partners
- Goal to bring in other agencies

the globus project™
www.globus.org

NSF Middleware Initiative R1

- Planned for March/April release date
- Globus 2.0+
  - MDS 2.2, GRAM 1.6, ...
- Condor-G
- NWS
- Enhanced packaging and testing
- Integration with campus directory services

Commercial Support

- Announced platform support from all major hardware vendors
- ISV support coming
  - Platform, Entropia
- Ability to purchase support contracts announced

Technology Futures

- Replica Location Service
- CAS Server
- Restricted delegation
- GRAM-II
- Accounting
- Web-services

the globus project™
www.globus.org

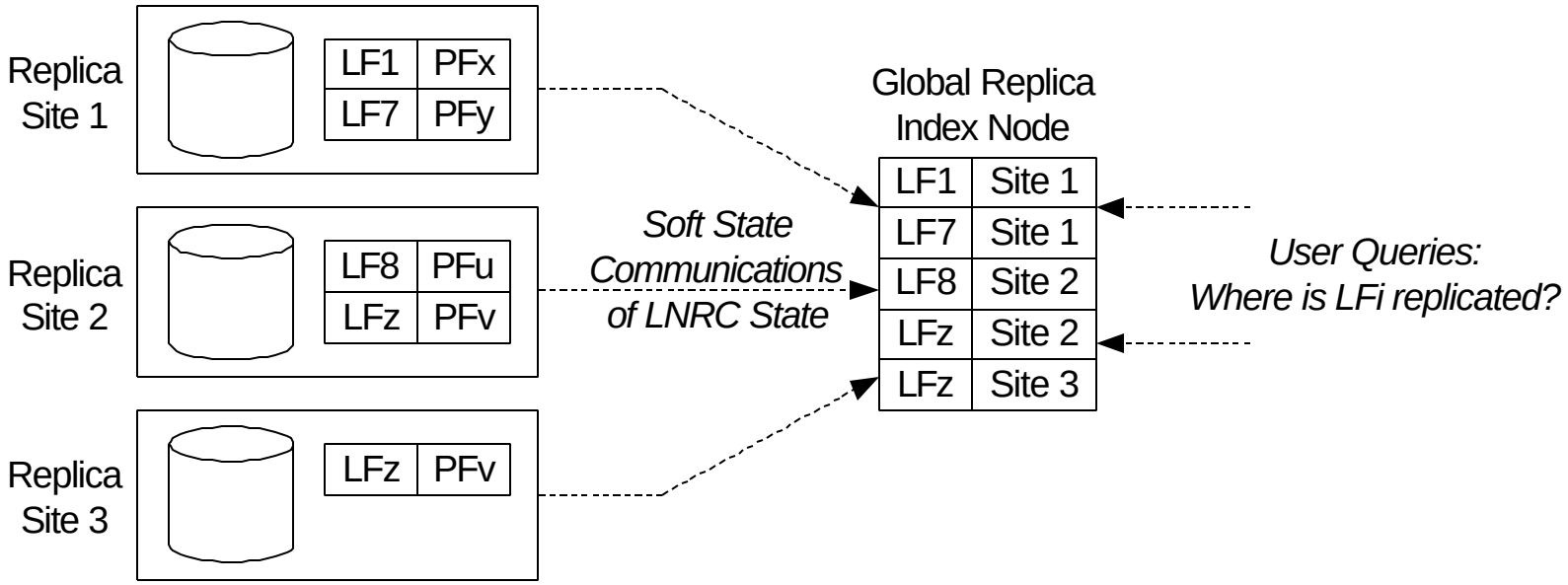
Replica Management

- Existing tools not currently part of Globus 2.0
- Plan to package and release replica management tools in Spring time frame
- Basic distributed catalogs may be supported in this release

A Replica Location Service Framework

- Applications may operate at different scales, have different resources and different tolerances to inconsistent RLS information
- We define a *flexible RLS framework*
- Allows users to make tradeoffs among:
  - consistency
  - space overhead
  - reliability
  - update costs
  - query costs
- By different combinations of 5 essential elements, the framework supports a variety of RLS designs

Example 1: A Centralized, Nonredundant Global Index

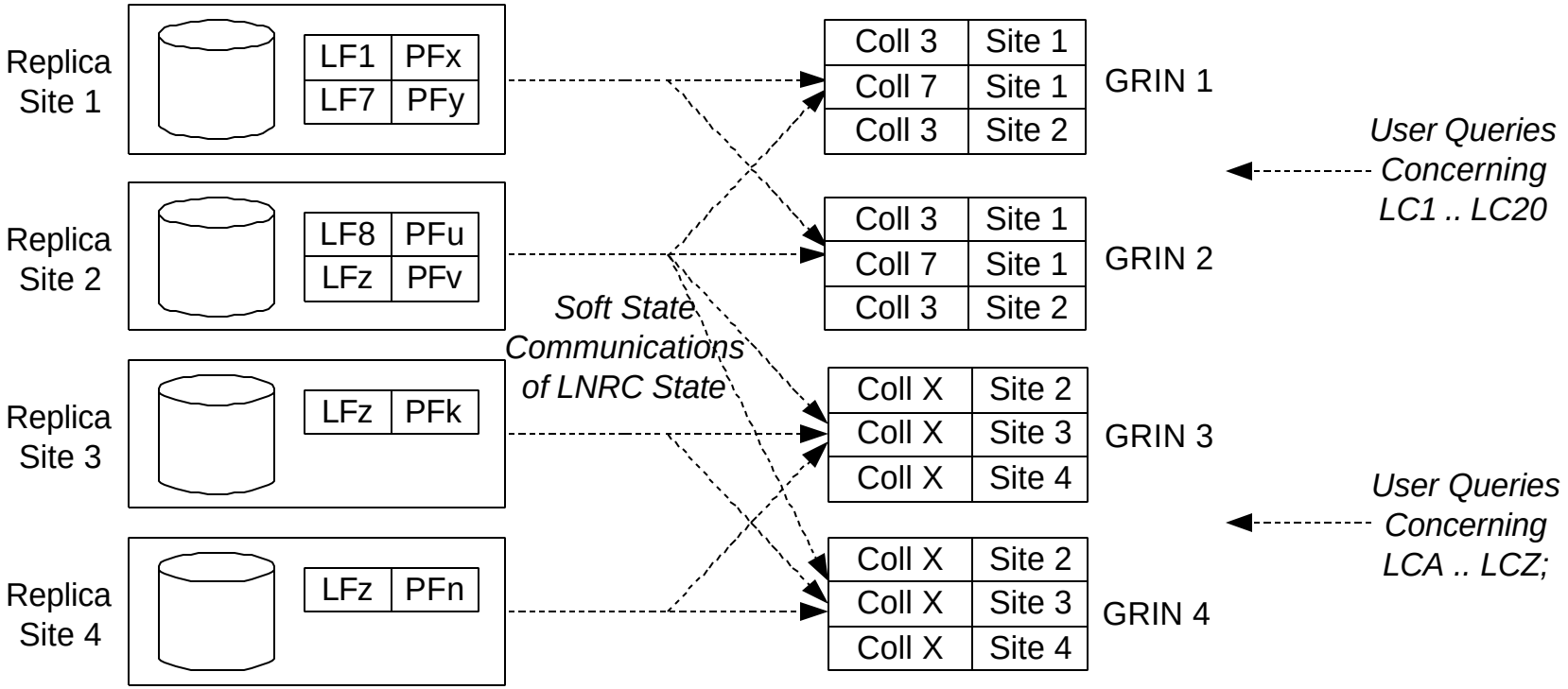


All updates sent to a centralized GRIN

Not scalable: All queries serviced by a single index

Not reliable: Single point of failure

Example 3: An RLS with Redundancy, Compression and Partitioning of Logical Collections



- Send collection information to GRINs (lossy)
- Advantage: Partition intelligently based on file contents, creation or access patterns

the globus project™
www.globus.org

Future MDS Activities

- MDS 2.2 in planning stages
  - Better performance
  - Better support for “pluggable” indexes and provider
  - Integrated event loop
    - Support for timed, and push-mode events
- Investigation into alternative index servers
  - Relational, XPath, ...

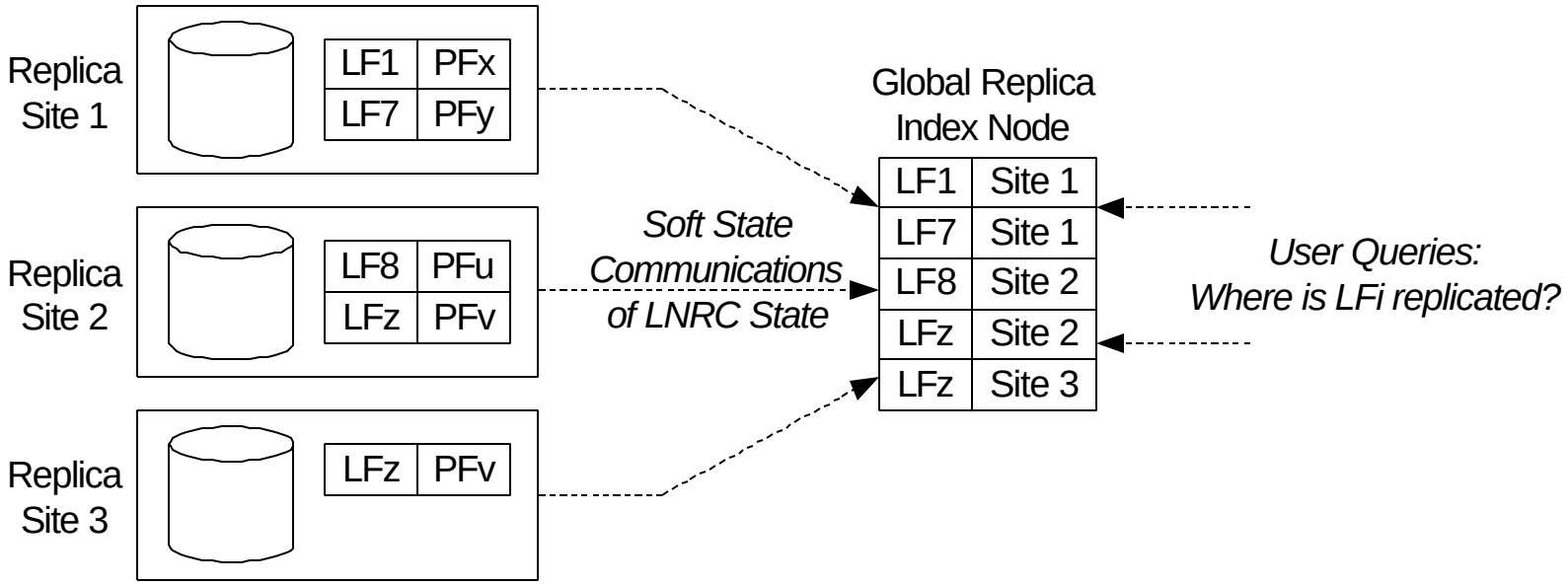
Motivation

- In a Data Grid, it may be desirable to create remote, read-only copies (replicas) of storage elements (files)
  - To reduce latency of data accesses
  - To increase robustness
- Need a mechanism for locating replicas
- ***Replica Location Problem:*** Given a unique logical identifier for data content, determine physical locations of one or more copies of that content
- ***Replica Location Service:*** a Data Grid component that maintains and provides access to information about physical locations of copies

A Replica Location Service Framework

- Applications may operate at different scales, have different resources and different tolerances to inconsistent RLS information
- We define a *flexible RLS framework*
- Allows users to make tradeoffs among:
  - consistency
  - space overhead
  - reliability
  - update costs
  - query costs
- By different combinations of 5 essential elements, the framework supports a variety of RLS designs

A Centralized, Nonredundant Global Index

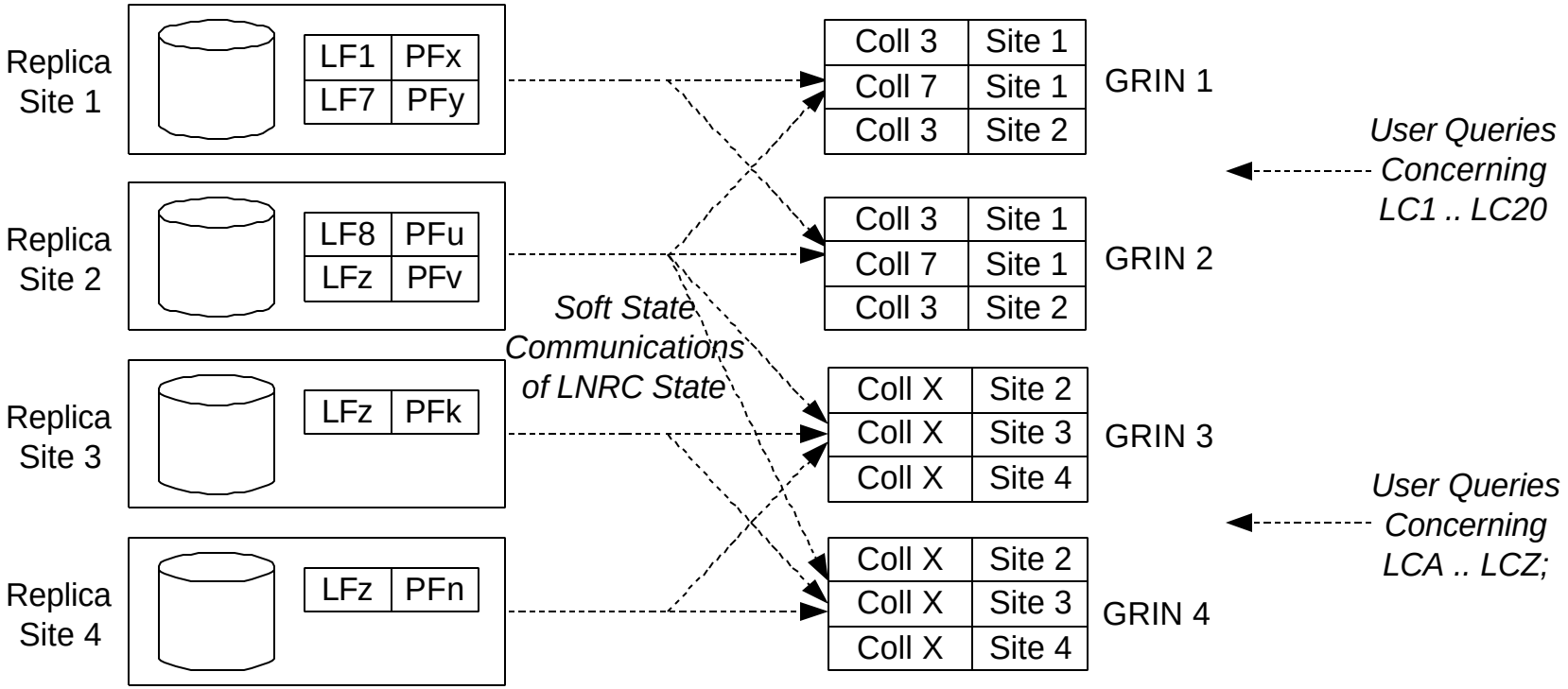


All updates sent to a centralized GRIN

Not scalable: All queries serviced by a single index

Not reliable: Single point of failure

Example 3: An RLS with Redundancy, Compression and Partitioning of Logical Collections



- Send collection information to GRINs (lossy)
- Advantage: Partition intelligently based on file contents, creation or access patterns

the globus project™
www.globus.org

GridFTP

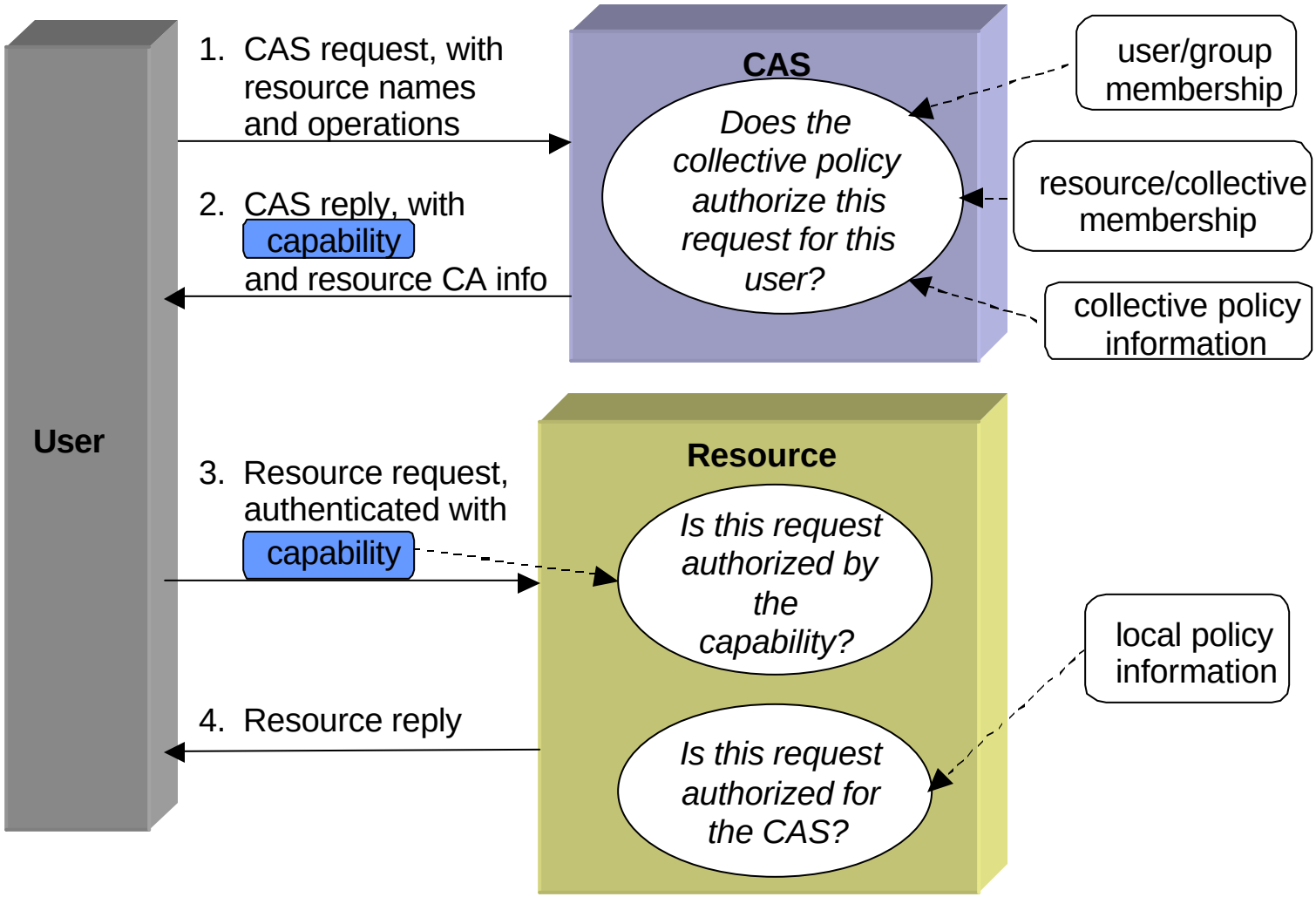
- Continue working on performance
- More robust striped server
- Integrated applications
  - E.g. Active mural
- Support for other platforms
  - HPSS, parallel file systems

Setting Community Policy

- Control of resources delegated to a community
  - E.g. community data sets, project allocations of compute cycles, ...
  - Resource “trusts” community administrator
- Community defines
  - Membership
  - Roles
  - Rights
- Need means for community to enforce community standards

the globus project™
www.globus.org

Community Authorization



Laura Pearlman, Steve Tuecke, Von Welch, others

the globus project™
www.globus.org

Gram 1.6

- Incremental changes to 1.5 protocol
- Support for advanced reservation on specific platforms
  - E.g. PBS
- Escape field in RSL to pass through schedule specific attributes

Resource Management Futures: GRAM-2

- Generic request protocol
  - Advance reservations
  - Multiple resource types: storage, networks, etc.
- Scoped/nested requests
  - Soft state for timeout and recovery
  - Nested transactions for complex requests
  - Policy evaluation points for restricted proxies
- Generic transport
  - SOAP/XML

Karl Czajkowski, Steve Tuecke, others

Grid Services

- Look at integration of Grid and Web Services architecture
  - Define basic, common underlying protocols
- Potential application to wide range of areas in business and science & technology
- Exploitation of significant commercial investment into tools and infrastructure